13 Rec'd PCT/PTO 28 NOV 2000
09/701525
WO 99/63708
PCT/IL99/00291

add A1

MODELING DATA SETS AND NETWORKS

FIELD OF THE INVENTION

The present invention relates to data sets and networks and in particular to ways of analyzing and configuring large data sets and networks.

BACKGROUND OF THE INVENTION

The development of computers and their introduction into almost every human activity has ushered in and catalyzed the proverbial "information explosion". Computers have enabled the assembly of enormous aggregations of information and the establishment of globe-circling information bulimic networks of people and machines. Computerized systems of all kinds generate deluges of data. The sheer size of these aggregations, networks and systems, hereinafter referred to as "information networks", and/or the rate at which they generate data often make them unwieldy and difficult to configure and manage.

For example, consider an Internet user using keywords to search the Internet for information. In response to a particular set of keywords, the Internet often presents such a user with a list of thousands and even hundreds of thousands of sites that may be able to provide the information the user desires. Even after the search is "sharpened" by modifying the original keywords with appropriate adjectives or by adding keywords, the proffered list of sites is often tediously and sometimes impossibly long.

Not only is the length of the list a frustration to the user, it also results in inefficient and wasteful use of Internet resources and contributes to "slowing down" the Internet. To the extent that the list is long, the user generally spends more time "mining" the list until he finds sites suitable to his needs. The longer the user, and other users like him, spend on the Internet searching for data, the more the communications capacity of the Internet is taxed and the longer it takes each user to access sites and download needed information. While the Internet seems to offer a cornucopia of unlimited information, the volume of the information offered often makes it difficult to access or use this information effectively.

Communication and command networks of interacting people and/or machines, common in today's business organizations, often present similar problems of information overload. In order to monitor and manage even a relatively simple network and optimize its performance, generally large quantities of data related to the performance of the machines and people in the network and the pattern of "information traffic" between them must be gathered and analyzed. To improve the efficiency of the network, or to adapt the network to changes in the tasks that it performs, the results of the analysis are applied to optimize or change the network configuration. The amount of data to be analyzed and the need to perform and apply

the analysis in time periods determined by events over which the network often has little influence puts a heavy strain on prior art methods for performing the analysis and applying the results of the analysis.

There is a need for improved methods for analyzing large and complex information networks and for configuring such information networks to improve the way they perform the tasks for which they are used.

SUMMARY OF THE INVENTION

It is an object of some preferred embodiments of the present invention to provide a method of modeling an information network.

One aspect of some preferred embodiments of the present invention relates to using the model for analyzing an information network. Preferably, model is used to detect and locate malfunctions in an information network. Alternatively or additionally, the model is used for configuring an information network. Alternatively or additionally, the model is applied to optimize the organization of a data base. Alternatively or additionally, the model is used to forecast how an information system will perform.

It is an object of some preferred embodiments of the present invention to provide a self configuring information network that learns from its own past functioning and adjusts and modifies itself in order to improve the efficiency with which it carries out the tasks for which it is used.

An information network comprises a set of network members that interact with each other and undergo changes when they interact. Each network member is characterized by a set of properties and is connected to other network members by various relationships. For example if the information network is an office network, the network members would be people and equipment in the office. A network member of the office that is a printer might be characterized by its printing speed and whether it prints in color or black and white. If the information network is a data base, the network members would be the different data elements in the data set.

Among the various types of relationships, hereinafter "connections", between members of an information network are physical, hierarchical and functional relationships. A cable connecting a computer to a printer is an example of physical connection. One person being a boss to another is an example of a hierarchical connection between two people. An example of a functional connection is a connection between a thermostat and an air conditioner whereby the thermostat turns the air conditioner on or off as a function of temperature that the thermostat senses.

The changes that occur in a network member of an information network and/or the work performed by the member are functions of the characteristics and features of the network member and changes that occur in other network members with which it is connected. Changes in a network member might also depend upon a change in an element external to the network that has a connection with the network member. An information network is said to be active when changes are occurring in its network members.

The set of all connections between network members in an information network is defined as a configuration of the information network. In accordance with a preferred embodiment of the present invention, a configuration of an information network comprises a structural configuration and a functional configuration.

The structural configuration is defined as the set of physical and hierarchical connections between network members. The physical connections, and generally the hierarchical connections, of an information network, are relatively static non-dynamic connections.

The functional configuration of an information network is the set of all functional connections between network members. The functional configuration may be considered to be a "dynamic configuration" of the information network that describes what network members do and how what one network member does is related to/affected by what other network members do. While the structural configuration of an information network is generally known and relatively easy to define and quantify the functional configuration is often very complex and difficult to define and quantify.

In accordance with a preferred embodiment of the present invention, a model of an information network is provided that provides a well defined quantifiable definition of a functional connection between network members of the information network and thereby a well defined quantified functional configuration of the network.

In some preferred embodiments of the present invention the functional configuration provided by the model can be used to analyze the information network and/or alert users and/or supervisors of the information network to malfunctions of the network. Alternatively or additionally, the functional configuration can be used to continuously and automatically adjust the structural configuration of the information network so as to optimize the performance of the information network or to adapt the information network to changes in the tasks that it performs. Information networks that use a functional configuration for continuous modification and optimization of the structural configuration of the information network may be considered self organizing autodidactic information networks. In a preferred embodiment of the invention,

the updating is performed relatively often, for example, every few seconds, minutes or days. Alternatively or additionally, the updating is performed periodically, such as once a month or a year. Alternatively or additionally, the updating is performed when, based on the determined functional connections, the activity of the information network is sub-optimal.

In accordance with a preferred embodiment of the present invention, a model of an information network that comprises a set of "nodes" that represent the network members of the information network. Each node represents a different one of the network members of the information network and is defined by at least one property that reflects the nature or characteristics of the network member that it represents. The nodes are connected to each other by relationships that mimic the relationships that connect network members of the information network and changes in nodes mimic changes in the members of the information network. As used herein, the term "nodes in an information network" should be taken to mean nodes in a model of the information network. Where elements of the information network are referred to, the term "members" is used exclusively.

In order to provide a well defined quantified functional configuration of an information network, in accordance with a preferred embodiment of the present invention, a measurable definition of a functional connection between nodes is defined. Two nodes are defined as having a functional connection when a change in one of the two nodes is connected to or correlated with a change in the other of the two nodes.

Nodes in an information network can be connected by different types of functional connections. For example, for a first task or activity of the information network two nodes might be functionally connected while for a second task or activity the same two nodes might not be functionally connected. In this case the first and second tasks may be considered to define two distinguishable types of functional connections.

Functional connections can also have different degrees of strength. For example, for a particular task or activity of an information network a first node might always be functionally connected to a second node but only sometimes connected to a third node. For the particular task or activity, the functional connection between the first and second nodes might be defined as stronger than the functional connection between the first and third nodes. Therefore, in accordance with a preferred embodiment of the present invention, nodes in an information network can be connected by different types of functional connections and functional connections between nodes can have different strengths.

In a preferred embodiment of the invention, the model comprises an activation network. Preferably, the learning of the model is event driven. When an event happens to a member in

the real world, a node, representing the member is activated. These events may be from outside the modeled network or they may be between members of the modeled network, both are termed herein external, as they are external to the model. In a preferred embodiment of the invention, the activation is propagated to other nodes of the model, based on functional connections between the node and the other nodes. After the activation spreads for a certain period of time and/or after a steady state is reached, the activation of activated nodes is correlated. This correlation may be temporally based. Alternatively or additionally, the correlation may be based on a known causative connection between the activations. The function used to test the correlation, may be a function of the external event, the node and other properties of the system. In a preferred embodiment of the invention, the temporal correlation may allow for a delay between the two activations. In a preferred embodiment of the invention, the delay is a window function. The window function, as with many other parameters of correlation, activation and external event treatment, may be a function of properties of the node, including a local memory, properties of neighboring nodes, properties of activated nodes and/or a type and/or properties of external event being analyzed. In a preferred embodiment of the invention, the window is used to model aspects of delay which may be expected in the real-world, for example, human response time, or mail delivery time. A functional relationship is then preferably updated based on the determined correlations. The updating may be a function of the above defined parameters and/or of any parameter and/or variable of the model. In a preferred embodiment of the invention, the updating is a function of whether the nodes at which the correlation was detected are both actors in a currently processed and/or related events.

In a preferred embodiment of the invention, the updating may create a functional connection between two nodes which were not previously connected. Alternatively or additionally, the update function may update existing connections.

In a preferred embodiment of the invention, the model is "harvested" and/or analyzed by applying one or mode inputs to the activation network and tracing the activation of networks as a result of these inputs. Thus, in some preferred embodiment of the invention, the updating may update any parameter of the activation network, including thresholds, weights, delays, forms of functions, decay and/or parameters of a node.

It should be appreciated that there might be no structural connection between the activated nodes. In addition, a node may become activated even if no event happened to its corresponding member. In a preferred embodiment of the invention, the activation is propagated as a time-varying signal. When the sum of arriving signals at a node is above a

threshold, the node is activated and/or generates an output signal, possibly at a delay. In a preferred embodiment of the invention, the threshold is a function of various properties of the node, parameters of functional connections to other nodes (such as weights in a graph representation), type of and properties of one or more external events which are being processed, whether the activation of the node is by external event or by an internal activation. In a preferred embodiment of the invention, the propagating activation is damped as a function of the distance from the originating activation. Alternatively or additionally, the output function of a node is depend on the distance from the event-activated node.

In a preferred embodiment of the invention, functional connections are modeled by weights between nodes in the model. In a preferred embodiment of the invention, when the functional relationship is updated, the weight is increased or decreased.

In a preferred embodiment of the invention, two activations are correlated based on the type of event which spawned the activations. In a preferred embodiment of the invention, only activations caused by a same type and/or a same group of event are correlated. Alternatively or additionally, the type of events to correlate are a function of the node for which correlation is being performed and/or is a function of other parameters of the model.

In a preferred embodiment of the invention, the activation of two nodes is correlated responsive to the propagation of activation in the model. In a preferred embodiment of the invention, nodes which are activated by an external event, are preferred for such correlation. In a preferred embodiment of the invention, only nodes which are activated by an external event are correlated. Alternatively or additionally, the weight and/or other parameters of the correlation and/or the updating function are dependent on whether the node become activated as a result of an external event and/or as a result of a propagating activation. In a preferred embodiment of the invention, two activations may be correlated even if one or both of them are not directly activated by an external event.

In a preferred embodiment of the invention, a node may have different thresholds for propagating an activation and for being activated to an extent that it partakes in a correlation.

In a preferred embodiment of the invention, the activation network is modeled using an architecture similar to that described in U.S. Provisional Patent application No. 60/057,818, titled "Heterogeneous Neural Network", filed September 4, 1997 by Yuval Baharav et al., now PCT application PCT/IL98/00430, the disclosure of which is incorporated herein by reference. In a particular example, each node is represented by one or more neurons. Different types of neurons and/or different parameters may be used for different node types, for example for nodes which represent users and for nodes which represent different types of resources. The

hierarchy of node types may be reflected by a hierarchy of neuron types. Rules which relate expected and/or allowed events and nodes are represented by non-learning connections. A typical learning rule for updating a weight between a node "i" and a node "j" (on a learning connection) can be $W_{ij}(new)=W_{ij}(1-\mu)*W_{ij}+\mu*a_i*a_j$. In some cases, data analysis neurons may also be provided for generating signals indicative of certain actions, such as certain rules being met. A more general analysis follows.

Let N_i represent the different nodes of an information network, where "i" is an integer index whose value indicates a particular one of the nodes. The set of all nodes in the information network is represented by $N = \{N_i\}$. Similarly let "FC_i" represent the different types of functional connections that connect nodes in the information network and $FC = \{FC_i\}$ the set of all different types of functional connections exhibited by the network. Classification of functional relationships can be defined by parameters of many different types, including, a time at which the event occurred, geography, state of the system being modeled and/or the members participating and/or properties of the members which participate in the functional interaction. A functional connection of the type FC_i between the "j-th" and "k-th" node can then be represented by $FC_i(N_j,N_k)$, where $FC_i(N_j,N_k)$ is assigned a value that represents the strength of the functional connection. For two nodes j and k that are not connected by a functional connection FC_i, $FC_i(N_j,N_k) = 0$. Using these symbols, the functional configuration of the information network is the set $\{FC_i(N_j,N_k)\text{: } FC_i \varepsilon FC;\ N_j \varepsilon N;\ N_k \varepsilon N\}$ of all functional connections $FC_i(N_j,N_k)$ that connect nodes in the information network.

A particular functional connection between two nodes is activated when a change in one of the nodes is correlated with a change in the other node as a result of the particular functional connection. Of the two correlated changes, wherein one of the changes is earlier than the other, the earlier change is considered to be a cause of the later change. A level of activation of the activated functional connection is defined as the magnitude of the earlier change times the strength of the functional connection. The activated functional connection is an output from the node in which the earlier change occurred and an input to the node in which the later change occurred.

Changes in a node in an information network, in accordance with preferred embodiments of the present invention, can depend upon inputs from other nodes, in different ways. In general a change in a node is a function of inputs from more than one node. In some cases the inputs to a node and changes in a node are represented by values of analogue functions. For example, a change in one node might be proportional to a continuous function of inputs from other nodes with which it has functional connections. In other cases changes in

nodes might be binary, *i.e.* they can only change from one to the other of two different states. Changes of state in a first node are the result of changes of state in other nodes that are communicated to the first node by functional connections that connect the first node to the other nodes.

Changes in a first node resulting from inputs from at least one second node are generally propagated by at least one output from the first node to at least one third node. Consider a first node having a functional connection with a second, third, fourth and fifth nodes. An output from the first node to the fifth node might depend on change in the first node that is a function of inputs from the second third and fourth nodes. The output to the fifth node is thereby a function, hereinafter referred to as a "transfer function", of the inputs to the first node. For example, the level of activation of the functional connection between the first and fifth node might be zero until the transfer function exceeds a threshold and thereafter be proportional to the value of the transfer function.

A transfer function is an algorithm by which a node processes inputs from a first at least one other node and provides at least one output to a second at least one other node. A node, in accordance with a preferred embodiment of the present invention, can comprise more than one transfer function. The transfer functions of nodes in a model of an information network, in accordance with a preferred embodiment of the present invention, are parts of the structural configuration of the information network.

In accordance with a preferred embodiment of the present invention the types and strengths of functional connections, *i.e.* the $FC_i(N_j,N_k)$ and their values, in an information network are defined as functions of correlations between changes that occur in nodes when the information network is active.

For each type of functional connection that an information network exhibits and/or that it is desired to investigate, a correlation test is defined. The correlation test for a particular type of functional connection is used to test if changes in different nodes of the information network are correlated with each other. When a change in one node is determined by the test to be correlated with a change in another node, then a "correlation event" has occurred between the two nodes. The correlation event is assumed to be the result of the two nodes being connected by the type of functional connection for which the correlation test is defined.

The correlation test for a type of functional connection can be a function of many different parameters and features of the information network. For example, the correlation test can depend upon a type of activity of the information network, properties of nodes, types of changes in nodes and time delays between the changes. In some preferred embodiments of the

present invention a correlation test provides a binary response, providing a "yes/no" answer as to whether two changes are correlated or not. In other preferred embodiments of the present invention the correlation test provides a numerical measure of degree of correlation between changes. In some preferred embodiments of the present invention the numerical measure can assume negative as well as positive values.

In accordance with a preferred embodiment of the present invention, each time a correlation event occurs between a first node N_j and a second node N_k, as determined by the correlation test for the functional connection FC_i, the value of the function $FC_i(N_j,N_k)$ is adjusted. $FC_i(N_j,N_k)$ can be adjusted in accordance with preferred embodiments of the present invention, in different ways. For example $FC_i(N_j,N_k)$ can be increased by a fixed amount every time an FC_i correlation event occurs between N_j and N_k. Alternatively, $FC_i(N_j,N_k)$ can be increased by an amount that decreases with increase in time separation between the correlated changes in N_j and N_k that produced the correlation event. $FC_i(N_j,N_k)$ might also be decreased if the time difference between correlated changes that produce a correlation event is greater than a certain time. Where the correlation test provides a numerical degree of correlation between changes, $FC_i(N_j,N_k)$ can be adjusted responsive to the value provided by the correlation test.

The transfer functions of a node in a network are chosen and adjusted so that that "output" correlation events of the node are correctly related to "input" correlation events, i.e. so that outputs from the node can be substantially accurately predicted from inputs to the node

Preferably, the functions $FC_i(N_j,N_k)$ are designed to decay in time so that if a particular functional connection $FC_i(N_j,N_k)$ between two nodes is not used, i.e. if no correlation events occur, the value of $FC_i(N_j,N_k)$ approaches zero and the functional connection atrophies. This assures that at any point in time the functional configuration of the information network is current. Different functional connections $FC_i(N_j,N_k)$ can be designed to decay to zero with different dependencies on time and different time constants. Time is measured in units relevant to the time scales and activities of the information network and advances only when the information network is in use.

In some preferred embodiments of the present invention functions $FC_i(N_j,N_k)$ and $TF_l(N_j)$ that are defined and determined for a model of an information network are used to analyze the network and/or alert users of the network to malfunctions of parts of the network. For example, models of information networks, in accordance with preferred embodiments of the present invention, can be used to identify bottle-necks in production processes, sources of

failures in computer networks and analyze the efficiency with which an organization accomplishes its tasks.

In other preferred embodiments of the present invention, functions $FC_i(N_j,N_k)$ and $TF_l(N_j)$ that are defined for an information network are used to continuously and automatically adjust the structural configuration of the information network so as to optimize the performance of the information network or to adapt the information network to changes in the tasks that it performs. This can be implemented relatively straightforwardly when parts of the structural configuration of the network comprise elements that can be adjusted under computer control. Changes that occur in the functions $FC_i(N_j,N_k)$ can be used by a computer to determine how to make adjustments of these elements. Information networks that use functions $FC_i(N_j,N_k)$, in accordance with a preferred embodiment of the present invention, to adjust and modify their own structural configurations in order to optimize performance or adapt to task changes are self organizing autodidactic information networks. For example a preferred embodiment of the present invention can be used to organize a data set to optimize data retrieval in response to the way users of the data set associate data in the data set. As the form of these associations change the data set can be automatically reorganized.

It should also be recognized that once functions $FC_i(N_j,N_k)$ for a model of an information network have been defined and evaluated and functions $TF_l(N_j)$ determined the model can be used to predict how the information network will react to various tasks or stimuli.

There is thus provided in accordance with a preferred embodiment of the invention, a method of modeling an information system having a structure, comprising:

detecting activations at at least two nodes of a structural model of the system;

correlating the detected activations; and

modifying at least one property of a functional relationship in a functional model of the system, responsive to the correlation.

Preferably, said correlating comprises correlating activations at nodes which are activated by an external event, responsive to said nodes being activated by a propagating activation in said model. Alternatively or additionally, at least one of said correlated activations is not directly caused by an external event in the system.

In a preferred embodiment of the invention, said property comprises a weight. Alternatively or additionally, said functional relationship is a direct relationship between said nodes. Additionally, said functional relationship does not directly relate either one of said nodes.

In a preferred embodiment of the invention, said activations are simultaneous. Alternatively, said activations are temporally overlapping. Alternatively, said activations do not temporally overlap.

In a preferred embodiment of the invention, the method comprises decaying a weight of said functional relationship responsive to a time since a last activation. Alternatively or additionally, said model is implemented using a neural network, in which each mode is represented by a neuron.

In a preferred embodiment of the invention, the method comprises modifying a structure of said information system using said modified functional model. Preferably, modifying a structure comprises optimizing a physical layout of said nodes. Alternatively or additionally, modifying a structure comprises optimizing a layout of communication lines between said nodes. Alternatively or additionally, modifying a structure comprises periodically harvesting said functional model. Alternatively, modifying a structure comprises continuously harvesting said functional model.

In a preferred embodiment of the invention, said information system is a computer network. Alternatively or additionally, at least one of said nodes represents a human being. Alternatively, said information system is a library.

In a preferred embodiment of the invention, said information system is a database.

In a preferred embodiment of the invention, the method comprises providing a permission to a real-world event responsive to said functional model. Alternatively or additionally, said information system is a data server and comprising using said functional model for enhancing data access. Alternatively, said information system is a distributed processing system and comprising using said function model for work allocation between elements of said processing system.

There is also provided in accordance with a preferred embodiment of the invention, a method of optimizing a data cache used in conjunction with a system, comprising:

determining a relation ship between events in said information system and access to data through said cache; and

modifying caching behavior of said cache responsive to said determination.

Preferably, determining a relationship comprises determining a functional model using a method as described above. Alternatively or additionally, said data cache comprises a file server. Alternatively, said data cache comprises a WWW site server. Alternatively, said data cache comprises a disk cache.

In a preferred embodiment of the invention, modifying caching behavior comprises selecting from a set of caching behaviors. Alternatively or additionally, modifying caching behavior comprises setting parameters for existing caching rules. Alternatively or additionally, modifying caching behavior comprises trading off between different classes of events in said system. Preferably, at least one of said classes of events represents a particular user of the system.

In a preferred embodiment of the invention, the method comprises reorganizing data in a data store cached by said cache.

BRIEF DESCRIPTION OF FIGURES

The invention will be more clearly understood by reference to the following description of preferred embodiments thereof read in conjunction with the figures attached hereto. In the figures identical structures, elements or parts which appear in more than one figure are labeled with the same numeral in all the figures in which they appear. The figures are listed below and:

Figs. 1A – 1C show schematically a structural configuration and two functional configurations of an office organization that are used to analyze the office organization in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

A modeling method in accordance with a preferred embodiment of the invention may be used for functional analysis of organizations, for example, for consulting purposes. Alternatively or additionally, the method may be used for identification of hidden centers of power and/or origins of failures. Alternatively or additionally, the modeling method may be used to model complex systems containing many elements, such as a traffic situation. Alternatively or additionally, the method may be used for identifying bottle-necks in a production process.

In a preferred embodiment of the invention, the models may be used to compare the behavior of a system to a model of the system to detect sudden changes from the norm. in one example, a sudden flurry of long-distance telephone calls may indicate a security problem with an employee.

In a preferred embodiment of the invention, the model is used for automatically generating rules, preferably based on the output of the model for a group of input sets. Alternatively or additionally, a model in accordance with a preferred embodiment of the invention, is used to analyze the response of a modeled system to a scenario, for example, a war.

It should be appreciated that a model in accordance with some preferred embodiments of the invention has a very high level of detail. Thus, the behavior of the modeled system under unexpected conditions may be more exactly modeled. In particular, a model in accordance with some preferred embodiments of the invention can model each and every member of a system, down to a low level, such a car in a country-wide traffic simulation. Usually, a model is designed analytically, with various simplifying assumptions. In preferred embodiments of the invention, few or no simplifying assumptions are made, at least with respect to the scale of the modeling.

A simple information network that is a sales office comprising salesmen and secretaries who communicate by e-mail can be used to illustrate definitions and functions used to model an information network in accordance with a preferred embodiment of the present invention.

The salesmen and secretaries, according to a preferred embodiment of the present invention, would be represented by nodes and a node would undergo a change every time "it" sent an e-mail message or read an e-mail message. For this simplified information network there might be only one type of functional connection of interest, a functional connection, $FC_o(N_j,N_k)$, representing "communication by email". A correlation event between two nodes would be a correlated "e-mail send" and "e-mail read". For two nodes exhibiting intense communication by e-mail $FC_o(N_j,N_k)$ would be relatively large while for two nodes exhibiting little e-mail communication $FC_o(N_j,N_k)$ would be relatively small.

A correlation function that would test for correlated "communication" changes in nodes would have no trouble telling which nodes were connected by an "e-mail send" or an "e-mail read" since each e-mail transmission would be identified by an address of a sender and receiver. However, the correlation function might return a numerical value for each correlated send and read, that decreases as the delay between the correlated send and read increases. For any delay greater than a certain amount, the correlation function might return a negative constant. Assume that for each correlated send and read for nodes N_j and $N_k$the value returned by the correlation function is added to $FC_o(N_j,N_k)$ and that between correlation events $FC_o(N_j,N_k)$ decays exponentially with a time constant of a day.

Given the above "scenario" it is highly probable that the best salesman can be identified with the node that has more and stronger connections $FC_o(N_j,N_k)$ to other nodes than any other node. The best salesman's node would be a center for a cluster of communicating nodes. Similarly, the more efficient, or better-looking, secretaries might be identifiable by nodes having numerous and strong connections to other nodes. The secretaries might delay as long as possible any responses to e-mail from a particularly ill-tempered salesman. The ill-tempered

salesman's node might be identifiable by the large number of negative connections. Finally, assume the best salesman has periodic bouts of depression that last a few days. The bouts of depression could probably be detected by an across the board decrease in the values of $FC_0(N_j,N_k)$ for communication connections between his node and other nodes.

Now assume that the sales office has a sales manager and a comptroller (represented by nodes). Assume that the sales manager handles very large sales that often carry a high risk of financial loss. As a result the sales manager works with a team of three "field salesmen" whose responsibilities are to gather financial and market information on each potential high risk sale. Company policy is that a decision to tender a sales proposal for a high risk sale requires the high risk sale receives a positive recommendation from the comptroller and from at least two field salesmen. From experience one of the three field salesmen is exceptionally capable and historically his recommendations have been very reliable. As a result, the sales manager takes a positive decision to submit a high risk sales proposal on the recommendation of this one field salesman alone and the comptroller as long as a second field salesman does not give a negative recommendation on the high risk sale.

Assume that in addition to the functional connection $FC_0(N_j,N_k)$ a "high risk e-mail" functional connection $FC_1(N_j,N_k)$ is defined. After tracking high risk e-mail with an appropriate correlation function, in accordance with a preferred embodiment of the present invention, it will of course be found that the nodes representing the sales manager, comptroller, and three field salesmen exhibit strong FC_1 connections between them. The sales manager's node will also have an FC_1 connection to a "high risk" secretary who handles the preparation and printing of high risk sales proposals.

Assume that a field salesman's recommendation in support of or against a high risk sale is represented by his node "e-mailing" a "+1" and "–1" respectively and that if he doesn't submit a recommendation at all his node doesn't activate his FC_1 connection with the sales manager's node. Similarly, assume the comptroller's input to the sales manager is represented by a 1 if he supports a high risk sale and a zero otherwise (including if he doesn't send a recommendation). Assuming the correlation function is appropriately defined to correlate with positive and negative decisions to submit a high risk sales proposal. Then, the relative strengths of the FC_1 connections from the salesman to the "capable" field salesman and the comptroller might have a value (after appropriate normalization) of 1 while FC_1 connections to the other field salesmen would have a value of 1/2. It will also be inferred that an appropriate transfer function that represents how the sales manager processes input from the other "high risk nodes" is a simple threshold test that requires that the sum of the inputs from the comptroller and the

field salesmen be greater than or equal to two. When this occurs there is a positive decision to submit a high risk sales proposal, the sales manager prepares a proposal and sends an output to the high risk secretary who prepares and prints the sales proposal.

Figs. 1A-1C show graphical representations of a structural configuration and two possible functional representations of an information network that is a small sales office for a printing business, in accordance with a preferred embodiment of the present invention.

The office has a sales manager, a secretary and a graphic artist. The sales manager is in charge of running the office and is boss to the secretary and graphic artist. The secretary is assigned responsibility for editing and printing sales proposals and letters composed by the sales manager and the graphic artist is in charge of preparing graphics that accompany sales proposals. The boss, secretary and graphic artist are connected by a LAN and additionally, the boss is connected by intercom to both the secretary and the graphic artist. The secretary's computer is connected to a black and white printer on which proposals and letters are printed. The graphic artist's computer is connected to a color printer on which graphics projects are printed.

Fig. 1A shows a graphical representation of a structural configuration 20 of the sales office, in accordance with a preferred embodiment of the present invention. The sales manager, secretary, graphic and printers are interacting network members of the information network and are represented by nodes in model 20. Circular nodes labeled respectively SM, SE and GA represent the sales manager, secretary and graphic artist. Square nodes labeled respectively BW and CP represent the black and white printer and the color printer. Wavy lines 22 between appropriate nodes represent the physical LAN connections and the connections between the printers and the computers. The intercom connection between the sales manager and the secretary and graphic artist are represented by broken wavy lines 24 between node SM and nodes SE and GA respectively.

Among the various types of interactions of the office personnel there are e-mail communications relating to graphics and e-mail communications regarding the editing and printing of proposals and letters. There are also graphics and editing transmissions to the printers. Hereinafter both graphics e-mail and graphics communications with printers are referred to as "graphics communications" and editing e-mail and editing communications with printers are referred to as "editing communications".

In accordance with a preferred embodiment of the present invention, graphics communications and editing communications define two types of functional connections, "$FC_G(N_j,N_k)$" and "$FC_E(N_j,N_k)$" respectively, between office personnel and/or equipment.

Every time a graphics e-mail or an editing e-mail is sent by a first one of the office personnel to a second one of the office personnel, and the second one of the office personnel reads the e-mail, a "graphics" or "editing" correlation event respectively occurs between the sender and reader. Similarly, a graphics or editing communication between one of the office personnel and a printer that starts the printer printing results in a graphics or editing event respectively. For simplicity, and clarity of presentation, time dependence of a correlation event on delay between sending and reading of an e-mail is ignored.

Figs 1B and 1C show graphically two possible functional configurations 30 and 40 respectively, for the sales office for functional connections $FC_G(N_j,N_k)$ and $FC_E(N_j,N_k)$.

Assume that when a graphics or editing correlation event occurs between two nodes a solid "graphics" line 26 or a dashed "editing" line 28 respectively is drawn between the nodes and that the number of lines between nodes is constantly being normalized to time in hours. At any one moment therefore, the number of graphics lines 26 and the number of editing lines 28 between two nodes in Figs 1B and 1C represents the average number of graphics communications and editing e-mail events occurring per hour between the nodes. The addition of a line between nodes for every e-mail event corresponds to adding a constant quantity to $FC_G(N_j,N_k)$ and $FC_E(N_j,N_k)$ every time a correlation event of their respective types occurs.

Other procedures for changing one of the functions, $FC_G(N_j,N_k)$ or $FC_E(N_j,N_k)$, as a function of a correlation event are possible and advantageous. Assume for example, it was desired to measure how long it takes to prepare graphics for a proposal and that projects were planned assuming a certain " planned delay" between a graphics project being assigned to the graphic artist and final graphics being printed. A correlation function that provided a weighted return having a maximum when a project was printed within a certain window of time centered on the planned delay could be useful. If the weighted return of the correlation function is added to $FC_G(N_j,N_k)$ for every graphics event, $FC_G(N_j,N_k)$ would be sensitive to the time it takes to produce graphics for a proposal.

Functional configuration 30 shown in Fig 1B is what might be expected if the sales office is running properly. Graphics lines 26 show that all graphics communications "moves" between the sales manager and the graphic artist and color printer. Editing lines 28 show that nearly all editing communications move between the sales manager, the secretary and the printer.

There are more graphics lines 26 between node SM and node GA than between node GA and node CP. This might be expected since it is reasonable that the sales manager and graphic artist communicate more frequently than graphic artist prints on the color printer.

Similarly there are more editing lines between node SM and node SE than between node SE and node BW. However, the ratio of the number of editing lines 28 between nodes SM and SE to the number of editing lines 28 between nodes SE and BW is not as great as the ratio of the number of graphics lines 26 between nodes SM and GA to the number of graphics lines 26 between nodes GA and CP. This also might be expected since editing and printing jobs would generally be smaller and more frequent than graphics printing jobs. The number of editing communications per editing printing job would generally be less than the number of graphics communications per graphics printing job. A low level of both graphics and editing communication is expected between the graphic artist and the secretary. A graphics line 26 and an editing line 28 between nodes GA and SE indicate this.

Functional configuration 40 shown in Fig 1C is what might be expected if the sales office has problems.

The secretary is a bit on the slow side and the graphic artist is bright and fast. As a result the sales manager prefers communicating with the graphic artist and very often asks the graphic artist to do the secretary's work of editing and printing letters and sales proposals. When the graphic artist prints a letter or a sales proposal the graphic artist usually does this on the secretary's printer that is much faster than the graphic artist's color printer. Because the graphic artist often performs the editing and printing tasks the graphic work suffers and sales proposals requiring graphic work often do not get out on time.

Functional configuration 40 makes the difference between the two office situations obvious. The shape of the functional configuration has changed noticeably. Functional configuration 40 is sharply skewed with respect to substantially symmetric functional configuration 30. Editing lines connect nodes SM and GA and nodes GA and BW. The secretary and graphics who communicated with each other in the "previous" office don't talk to each other at all. There are no graphics or editing lines between nodes GA and SE.

Another example illustrates how a preferred embodiment of the present invention can be applied to provide a self organizing data base.

Consider an information network that is a large computerized document library in which documents can be searched for and located using keywords and from which they can then be down loaded. The library in effect, is a large data base stored in a computer memory, which data base comprises groups of keywords that represent documents and individual keywords or groups of keywords that are used in searches for documents.

In accordance with a preferred embodiment of the present invention, keywords and groups of keywords used in searching for documents and groups of keywords used in defining documents are nodes in a model of the library.

The structural configuration of the library comprises the way the keyword nodes and document nodes are located or stored with respect to each other in the library memory, i.e. the relationships between the addresses of keyword nodes and document nodes in the computer memory housing the library data base.

A correlation event occurs between a keyword node and a document node if, after querying the library with the keyword represented by the keyword node, a user accesses or downloads the document represented by the document node. A correlation event occurs between the document node and a second document node if a reference in the first document causes the user to reference the document represented by the second document node. The correlation events between keyword nodes and document nodes are registered by appropriately defined functions $FC_i(N_j,N_k)$. For nodes representing keywords and documents that are frequently and repeatedly referenced together $FC_i(N_j,N_k)$ will be large.

In accordance with a preferred embodiment of the present invention the values of $FC_i(N_j,N_k)$ are periodically automatically reviewed. Following each review the library memory is automatically reorganized so that key word nodes and document nodes for which $FC_i(N_j,N_k)$ is large are rapidly associated together and located when the library is searched for information that the documents contain.

Such a library is a self organizing data base that learns from experience which data items are related, how strongly they are related, and then groups related data items "close" to each other in memory. Eventually the library memory will be organized into clusters of related keywords and documents that might for example be located in the same or nearby blocks of memory in the library or might be members of a linked data set.

The clusters of related data items of course reflect the way users of the data base associate items in the data base. If the users should change the way they associate data items in the library, the library will recognize the change because the values of the functions $FC_i(N_j,N_k)$, in accordance with a preferred embodiment of the present invention, will change in response to the new way data items are associated. The library will then reorganize itself into a new pattern of clusters to reflect the new values of the functions $FC_i(N_j,N_k)$. The library can learn and adapt itself to change.

In a preferred embodiment of the invention, the changes in the model are applied to the real-world library database, at the end of every day. Alternatively or additionally, these changes

are applied at the end of every search and/or data entry. In a preferred embodiment of the invention, searches performed by a faculty member will have a significantly greater effect on modifying functional connections in the model than will those of a student.

Another example illustrates the use of a preferred embodiment of the present invention as a prognostic or forecasting tool in a medical application, in which the invention is used to determine relationships between symptoms and measured physiological parameters.

Many adult males suffer from a sleep disturbance phenomenon called apnea. Apnea involves instances of breathing cessation that cause a sufferer to wake up numerous times during a night and not only leaves a person tired but can result in serious damage to the body and even death.

In order to understand apnea and predict which body changes or confluence of changes during sleep trigger an occurrence of apnea a patient might be fitted with sensors that measure different parameters of his body functions while he sleeps. For example, he might be fitted with sensors that track body temperature, blood pressure, heart rate, respiratory rate, rapid eye motion and brain waves, and a pickup microphone to register the sounds of his snoring.

Each sensor is represented by a node. Functional connections $FC_i(N_j,N_k)$ between nodes are established as a result of correlations between changes in measurements of the various sensors. For example, it might be found that periods of rapid eye motion precede by a certain period of time a sudden rise in blood pressure or heart rate and that this is then followed by an arrhythmia event and a sudden small dip in blood pressure. These and other events might correlate with the onset and severity of an apnea event as monitored by snoring sounds that the patient makes. Once the functional connections $FC_i(N_j,N_k)$ are evaluated and transfer functions inferred for the various nodes, a functional configuration of apnea events results, in accordance with a preferred embodiment of the present invention, that might be used to clarify how they are triggered and how they might be prevented.

In a preferred embodiment of the invention, a model is made of communication networks, for example, telephone networks and/or computer networks. As a result of the model, it is possible to determine which geographical locations have heavier telephone traffic and at what time. In a preferred embodiment of the invention, external events include news events, vacation schedules, television schedules and other happenings which affect a daily schedule of many people. In a preferred embodiment of the invention, the nodes of the network may represent countries, cities, local interchanges, streets and even individual subscribers.

In a preferred embodiment of the invention, the above modeling method is used for optimizing the location of files on a disk. In one example, when a "mega application" is loaded

in the Windows95 operating system, a large plurality of DLL files are loaded. Typically, these files are not located in a physically near location, so their loading takes a long time. In a preferred embodiment of the invention, the above described modeling method is used to analyze which DLLs are loaded at the same time and/or in response to loading the same programs. Thereafter, the physical and/or logical location of these files may be changed to reflect the way a particular user uses his machine.

In another embodiment of the invention, the above modeling method is used for optimizing data retrieval, for example in caches and data servers. In a preferred embodiment of the invention, the above modeling method and/or other, known, modeling methods, are used to determine relationships between data requests and events accepted by a system which generates these requests. In one example, events are correlated with sequences of disk blocks being read. In another example, the request for a particular WWW page from a server, by a particular user is correlated with other page requests by the user, to determine expected pages to be read. In a similar example, a file server may read ahead and/or send ahead files which, based on a modeling of the outside system, appear to be likely to be read. Thus, the decision whether to read data into a cache and/or what "grade" to assign data in a cache may be related to external event and/or to sequences of block reads. These considerations may be applied both to read caches and to write caches.

In some cases, a particular event may be related to a set of relationships between blocks. For example, in a microprocessor, an address look ahead cache (which retrieves instructions which may be required in future machine cycles), can be optimized for a particular program and/or instance of a program execution. In one example, the above modeling method is used to determine relationships between conditional branchings and events. This data may be used to generate a more optimal cache-rule table, which table is downloaded to the cache. In some cases, a simulation of the program may be used instead of a real-life execution. Alternatively or additionally, to complete cache instruction rules generated by modeling, the modeling may be used for selecting a particular cache rule set, from a set of available rules. As indicated above with respect to a WWW server, the relationships may be associated with a particular user, IP address, program, source WWW site, time of date and/or other parameters of the event. It is noted that a plurality of users may be accessing a cache (e.g., of a WWW server, file server, disk, CPU) at the same time. Various tradeoffs may be used, for example based on available cache space or based on the expected cache requirements. In a system including several caches the caches may optionally communicate and/or otherwise be synchronized with respect to their caching behavior.

In another example, the above modeling method is used for planning work schedules and/or dividing-up work between actors, based on a modeled relationship of delaying and interaction between actors. These actors may be, for example, computer programs (for example in the case of distributed computing) or people, for example sub-contractors in a building project. In a particular case of work division, sub-processes may be distributed between processors based on an expected (from a model) amount of communication between particular sub-processes.

In a preferred embodiment of the invention, the above modeling method is used to provide a security system for a computer system and/or network. A network may be described as a set of users and a set of resources (e.g., files, database items, communication ports and network devices). Each resource and each user are represented by one or more nodes. Events occurring in the network are audited and used as training inputs. In a preferred embodiment of the invention, the system adapts to these events by changing the weight, delay function and/or other parameters (as described above) of the neurons and/or their connections. Thus, the model can learn to reflect the functional relations in the system. Preferably, the learning is event driven. Alternatively or additionally, the learning is sampling driven, for example by periodically sampling events. Alternatively or additionally, the learning is statistical, by taking in to account only some of the events in the system.

In a preferred embodiment of the invention, when a user node is activated by an external event – such as a user accessing a file, the connection between the node representing the user and the node representing the resource is changed according to a correlation function. The correlation function may be temporally and/or node properties based. A non-active connection may decrease with time according to the system's decay parameter.

After an initial training period, the system reaches a quasi-steady state, in which the reflection (of the system by the model) suffices. The reflection is a densely inter-linked database, on which a clustering method may be applied on, to obtain usage profiles.

A clustering algorithm can yield a normal usage profile, from which the un-likelihood of an action (A user tries to access a resource in a certain mode and parameters) can be derived. These norm profiles are preferably stored in a second database.

When an action is executed in the network permission is requested from the system. The system derives the un-likelihood of this action, and compares it to pre-defined thresholds, thus taking the response decision. The thresholds are defined according to the security level assigned to the resource. Threshold decision is preferably determined by a trade-off between the twin dangers of misuse and false alarms. In some cases, the permission is

granted n a case-by case basis, in other cases, the security system can generate estimates of un-likelihood based on a pattern of actions by a particular user and/or programs executed, written and/or spawned by the user. In some embodiments, the model is simultaneously utilized in two manners, a first manner in which the model learns the system activity so that it can be harvested and a second manner in which the model mimics the system activity and generates a signal if an unlikely event occurs.

Each action can also serve as an additional event, learned by the system. The adaptation process is preferably designed in such a way that the latest events have more influence than old ones. In this way the system tracks trends. The "forgetting factor" is preferably set automatically, according to the network stationarity time-constant.

Both a computer system (which includes a plurality of "user" programs and a plurality of resources on a single computer and a computer network in which the resources and/or the users are more distributed, can be modeled using the above method. in a particular example, the above method is used to monitor a LAN system for detecting hacking in from an outside computer or by a disgruntled worker on the same LAN. In another particular example, the above system can detect computer virus-like behavior by detecting undesirable (which can be trained into the system), disallowed and/or unlikely activities by a particular program or a set of programs.

It will be appreciated that the above described methods of applying modeling may be varied in many ways, including, changing the order of steps and which steps are performed on-line and which offline. In addition, a multiplicity of various features, both of method and of devices have been described. It should be appreciated that different features may be combined in different ways. In particular, not all the features shown above in a particular embodiment are necessary in every similar preferred embodiment of the invention. Further, combinations of the above features are also considered to be within the scope of some preferred embodiments of the invention. Also within the scope of the invention are computer readable media, such as diskettes, which include software, which when installed on a computer form a machine capable of modeling, as described above. Additionally, although the above invention has been described mainly as a method, a computer including software and/or other hardware suitable for carrying out the method is also in the scope of the present invention. Such a computer, hardware and/or software may be distributed. When used in the following claims, the terms "comprises", "includes", "have " and their conjugates mean "including but not limited to".

Variations of the above-described preferred embodiments will occur to persons of the art. The above detailed descriptions are provided by way of example and are not meant to limit the scope of the invention, which is limited only by the following claims.

09701535 .012301